R E P O R T R E S U M E S

ED 017 957 72 CG 001 796

A BATTERY OF TESTS ON GENERAL EDUCATIONAL DEVELOPMENT FOR POST-ELEMENTARY SCHOOLS. VOL. I.

BY- BENTWICH, J. AND OTHERS
HENRIETTA SZOLD INSTITUTE, JERUSALEM (ISRAEL)

REPORT NUMBER HSI-RR-NO-112 PUB DATE MAR 67
RUTH BRESSLER CENTER FOR RESEARCH IN EDUCATION
CONTRACT OEC-4-21-011
EDRS PRICE MF-$0.25 HC-$0.96 22P.

DESCRIPTORS- *TEST CONSTRUCTION, TEST RELIABILITY, TEST VALIDITY, *HEBREW, *ACHIEVEMENT TESTS,

THIS TEST BATTERY IS DESIGNED TO BE USED AS AN AID IN COUNSELING AND GUIDANCE FOR PUPILS IN THE NINTH AND TENTH GRADES OF ACADEMIC HIGH SCHOOLS IN ISRAEL. AS THE INTENT IS TO MEASURE THE PUPIL'S ABILITY TO DO CRITICAL THINKING IN BROAD AREAS, THE TEST BATTERY MEASURES GENERAL EDUCATIONAL DEVELOPMENT RATHER THAN SPECIFIC ACHIEVEMENT SKILLS IN THE AREAS OF MATHEMATICS, SCIENCE, READING COMPREHENSION, SOCIAL STUDIES, AND ENGLISH LANGUAGE. THE STATISTICAL INFORMATION CONCERNING RELIABILITY, VALIDITY, INTERCORRELATIONS, AND NORMS ARE BASED ON A NATIONAL SAMPLE OF NINTH AND TENTH GRADERS IN ACCREDITED ACADEMIC HIGH SCHOOLS. INSTRUCTIONS FOR ADMINISTERING THE TESTS ARE ALSO GIVEN. THOUGH THE BATTERY IS TO BE USED IN COUNSELING, IT IS EMPHASIZED THAT IT IS NOT YET COMPLETE AND IS INADEQUATE IN ITS PRESENT FORM TO FUNCTION AS A BASIS FOR THE REMOVAL OF A PUPIL FROM AN ACADEMIC HIGH SCHOOL TO A NONACADEMIC ONE. THE NORMS ALSO NEED TO BE EXTENDED TO COVER VOCATIONAL SCHOOLS AND AGRICULTURE SCHOOLS. (CG)

ED017957

THE HENRIETTA SZOLD INSTITUTE

NATIONAL INSTITUTE FOR RESEARCH
IN THE BEHAVIORAL SCIENCES

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

NO BR no.
PA 72
Contract no.
OE 4-21-011

A BATTERY OF TESTS

ON GENERAL EDUCATIONAL DEVELOPMENT

FOR POST-ELEMENTARY SCHOOLS

by

J. BENTWICH, J. LEVIN, D. ORMIAN

RESEARCH REPORT NO. 112

PUBLICATION NO. 456

MARCH 1967

CG 001 796

THE HENRIETTA SZOLD INSTITUTE

National Institute for Research in the Behavioral Sciences

A BATTERY OF TESTS OF EDUCATIONAL DEVELOPMENT

by

J. Bentwich, J. Levin, D. Ormian.

Project on Construction of a set of Instruments in order to Assess Achievements, Aptitudes and Attitudes of Students in Academic Secondary Schools in Israel.

This project was made possible by a grant from the United States Government, Department of Education. Project No. OE-4-21-011.

TABLE OF CONTENTS

		Page
A.	DEVELOPMENT OF THE PROJECT AND DESCRIPTION OF THE BATTERY	2 3
B.	STATISTICAL PROPERTIES The Sample, Reliability, Validity, Intercorrelation, Time-span.	4
C.	APPLICATIONS OF THE BATTERY	10
D.	DISTRIBUTION AND NORMS	13
E.	INSTRUCTIONS FOR ADMINISTERING THE BATTERY	17
F.	REFERENCES	19

LIST OF TABLES

TABLE I.	Reliability of the subtests and of the total battery.	5
TABLE II.	Correlations of subtest scores with the total scores.	6
TABLE III.	Intercorrelations, means and standard deviations of the subtests for grade 9.	8
TABLE IV.	Intercorrelations, means and standard deviations of the subtests for grade 10.	9
TABLE V.	Interquartile ranges of the distribution of class means of each subtest for grade 9.	12
TABLE VI.	Interquartile ranges of the distribution of class means of each subtest for grade 10.	13
TABLE VII.	Number of points necessary to add or subtract to scores of pupils participating in 5 subtests.	15
TABLE VIII	Norms and percentile scores of the mean score.	16

LIST OF FIGURES

A.	Distribution of scores (average of 6 subtests in grades 9 and 10).	14 a

A. Development of the Project.

The initial plan of the project was prompted by a series of research publications on current examinations and methods of assessment in high schools. These studies emphasized the inadequacies of the evaluation methods with respect to reliability, distribution of scores and standardization. Surveys of school failures and drop-outs, published at about the same time, showed serious bottlenecks in the secondary school system (see list of references on page 10).

There is a close relation between these issues, since success and failure in school, among other things, depend on the evaluation methods employed by the teachers, and on the availability of guidance aided by proper instruments.

Adequate assessment methods are also essential for adaptation of curriculum to the level of students and their rate of progress. Since the structure of assessment instruments should conform to these purported objectives, it is important to keep one main objective in the foreground. There is consensus of opinion among educational experts emphasizing guidance as the most urgent task.

Since standardized intelligence tests are already available in Hebrew, it was decided to focus on educational attainment. Owing to the diversity of curricula in the lower grades of the high school, it seemed preferable to construct a test battery to measure general educational development rather than strict achievement tests. Also, from the guidance point of view, the general level of a student is far more important than specific items of information.

The well-known batteries I T E D (Iowa Tests of Educational Development) and S T E P (Sequential Tests of Educational Progress) served as a model for the present battery:

"...THE TESTS ARE DESIGNED TO MEASURE MUCH MORE THE GENERALIZED SKILLS. THEY ARE INTENDED TO MEASURE THE PUPIL'S ABILITY TO DO CRITICAL THINKING IN THE BROAD AREAS DESIGNATED; THEY ARE CONCERNED

NOT SO MUCH WITH WHAT THE PUPIL HAS LEARNED, IN THE SENSE OF SPECIFIC INFORMATION, BUT RATHER WITH HOW WELL HE CAN USE WHATEVER HE HAS LEARNED IN ACQUIRING, INTERPRETING AND EVALUATING NEW IDEAS, IN RELATING NEW IDEAS TO OLD, AND IN APPLYING BROAD CONCEPTS AND GENERALIZATIONS TO NEW SITUATIONS ON THE SOLUTION OF NEW PROBLEMS. THESE ARE SOME OF THE OUTCOMES NOT ONLY OF AN EFFECTIVE COURSE OF FORMAL SCHOOL INSTRUCTION, BUT ALSO OF ANY OTHER GENUINELY EDUCATIONAL EXPERIENCE, WHETHER FORMAL OR INFORMAL, DIRECT OR INCIDENTAL, IN-SCHOOL OR OUT-OF-SCHOOL..."
(From Manual for Teachers and Counsellors, I T E D.)

Description of the Battery.

The battery comprises 6 subtests:

1) Mathematics
2) Science (Physics, Chemistry, Biology)
3) Reading Comprehension - literature.
4) Reading Comprehension - social studies.
5) Social Studies - general information.
6) English Language.

The tests differ in their relative dependence on the school curriculum material. In the Sciences (1 and 2 above) and English Language (6) the connection is relatively close. While in the humanistic studies (3, 4 and 5) connection to the school material is, relatively, loose and the factors of general knowledge and the comprehension of basic concepts are pre-eminently emphasized. Clearly the weight of emphasis on school material, on the one hand, and on general knowledge, on the other, is not equal in every item of the test.

Tests included in the battery:

1) Mathematics - 24 questions in Geometry and Algebra.
2) Science - 33 questions in Physics, Chemistry and Biology.

The questions are related to the material contained within the school curriculum, however a pupil's score depends also on the extent of his extracurricular reading.

3) Reading Comprehension - literature*. The test examines comprehension of extracts culled from literature. Six extracts are given, each followed by a number of questions. The total number of questions is 20.

4) Reading Comprehension - social studies. The test includes 5 extracts from social studies with a total of 22 questions.

5) Social Studies - general information. 22 questions on History, Jewish History, Economics, Civics and Geography. The questions are, in part, drawn from the school curriculum material while the rest depend largely on the general knowledge of the pupill.

6) English Language. 55 questions covering vocabulary, grammar and reading comprehension.

The test items are arranged in order of difficulty.

B. Statistical Properties.

1. The Sample.
2. Reliability.
3. Validity.
4. Intercorrelations.
5. Time-span.

1. The Sample.

The sample used as a basis for the computation of norms is a national sample of the 9th and 10th grades of accredited academic high schools. The statistical data reported here are based on this sample.

*Literary extracts taken from:

The Haggadah; S.H. Bergman; L. Goldberg; A. Kovner; G. Shofman; Y. Shœnberg.

The sample is stratified and from each stratum a random sample is chosen. The stratification was based on:

(i) Type of settlement wherein the school is placed (big cities, various sized towns and other kinds of settlements)

(ii) The scholastic level of the school. Since there is no accepted uniform measurement of the scholastic standard of a school, we used the percentage of successful matriculation candidates in the preceding year, 1965.

The sample comprises 370 pupils in the 9th grade (11 classes) and 370 pupils in the 10th grade (13 classes).

The tests were given during the last trimester of the 1966 school year.

2. Reliability.

Since no parallel test is available, we employed the split half method of computation of reliability, dividing each test into odd and even items. The results are given in Table I.

TABLE I.

Reliability of the Subtests and of the Total Battery*

Test	Grade 9	Grade 10
Mathematics	.75	.72
Science	.68	.71
Reading Comprehension - literature	.57	.65
Reading Comprehension - social studies	.53	.60
Social Studies - general information	.69	.66
English Language	.85	.87
Total Battery	.90	.91

The highest reliability is in the English Language test (.85 - .87). This is, possibly, because it is the longest test and reliability increases with length. The reliabilities of the tests in the Sciences (.68 - .75) are higher than those in the humanistic studies (.53 - .69).

*Corrected by the Spearman-Brown Formula.

The reliability of the subtests is not high enough for a construction of profiles of the individual student. This is due to their brevity and heterogeneity of content. Higher reliability could have been obtained had the subtests been constructed of a homogeneous pool of items; however, this would confute the purpose of global measurement. We considered it preferable to include a great number of areas at the expense of reliability. Reliability could be raised by increase of length, but this would require a longer time span of testing, which would restrict the applicability of the battery.

Use of the mean score of the subtests is therefore recommended. It is based on 176 items. Its coverage is very broad and the reliability (.90) is high enough for individual assessment. (Mean score instead of total score is employed, to allow for cases of omission of a subtest).

Employment of raw scores weighs the tests unequally, but the departure from equality is small as can be seen from Table II.

TABLE II.

Correlations of the Subtest Scores with the Total Scores

Subtest	Grade 9	Grade 10
Mathematics	.68	.65
Science	.69	.71
Reading Comprehension - literature	.60	.63
Reading Comprehension - social studies	.65	.68
Social Studies - general information	.69	.59
English Language	.75	.76

The range of the relative weights, i.e., the ratios of the correlations is between 1.0 and 1.3. The tests may be considered as equally weighted. The largest weight is of the English test which is only slightly larger than the others (1.3 in both grades). This slight difference does not justify the additional work involved in application of weighted averages.

3. Validity.

Methods of determining validity are quite complex since there is rarely a single, unequivocal criterion by which to measure validity. In fact, validity may be determined only on the basis of protracted research and systematic exploration of relations between the tests and practical and theoretical criteria.

The following evaluation is based on data collected within the scope of the present project. Further information will be collected in due course of time in the context of a follow-up study.

The content validity of the battery was established by judgment of teachers, school supervisors, and experts on the subject matter of each subtest respectively.

Concurrent validity with respect to academic achievement is based on correlations between the battery score and point-grade averages of 21 classes. The median correlation is .50. The interquartile range is .32. These values underestimate the validity because of the restriction of range in each class. This becomes evident when we note the large difference in scores among schools. However, computation of the correlation between the battery scores and school grades for the whole sample would involve standardization of school grades on one consistent scale.

The validity with respect to prediction of academic achievement, success and failure in school, dropping-out and effectiveness of the tests for adaptive treatment and placement purposes, requires a follow-up study.

4. Intercorrelations.

The intercorrelations, means and standard deviations of the sub-tests are given in Tables III and IV. The entries in the diagonal are the split-half reliabilities corrected by the Spearman-Brown Formula. The entries above the diagonal are the raw correlations. Below the diagonal are the correlations corrected for attenuation.

TABLE III.

Intercorrelations, means and standard deviations of the subtests for grade 9.

(N = 370)

Subtest	1 Mathematics	2 Science	3 Reading Comp. - Literature	4 Reading Comp. - social studies	5 Social Studies - general information	6 English Language
1 Mathematics	(.75)	.52	.27	.30	.45	.32
2 Science	.73	(.68)	.29	.37	.47	.29
3 Reading Comprehension-literature	.42	.47	(.57)	.47	.37	.35
4 Reading Comprehension-social studies	.48	.62	.85	(.53)	.41	.37
5. Social studies-general information	.63	.68	.59	.68	(.69)	.28
6 English Language	.40	.38	.50	.55	.36	(.85)
Mean	13.6	17.5	13.6	13.6	12.6	33.4
Standard Deviation	4.0	4.3	2.8	3.1	4.3	7.8

TABLE IV.

Intercorrelations, means and standard deviations of the subtests for grade 10.

(N = 370)

Subtest	1 Mathematics	2 Science	3 Reading Comp. - literature	4 Reading Comp. - social studies	5 Social Studies - general information	6 English Language
1 Mathematics	(.72)	.49	.31	.36	.37	.25
2 Science	.68	(.71)	.29	.37	.32	.38
3 Reading Comprehension - literature	.46	.43	(.65)	.51	.36	.35
4 Reading Comprehension - social studies	.55	.57	.82	(.60)	.36	.40
5 Social Studies - general information	.54	.47	.55	.57	(.66)	.21
6 English Language	.32	.48	.47	.56	.28	(.87)
Mean	15.3	19.0	14.0	14.3	13.4	39.9
Standard Deviation	4.0	4.8	3.2	3.0	4.0	8.0

The highest correlations (see correlations below the diagonal) are between:

1) Reading Comprehension tests - literature and social studies (.82-.85). This is due to the similarity of function and content.

2) Mathematics and Science (.68-.73). The two science disciplines in the battery.

Other correlations are lower and the lowest are found in English; however, a thorough investigation of the composition of the battery would require a factor analytic study.

5. Time-span.

Part E gives details of the times allowed for each test. These time-spans were determined in the pre-test. It is customary to allow 85% of the examinees to finish the test. This, too, was our basis for deciding on the time-span for each subtest.

The tests, therefore, depend little on speed of performance. It has been found that additional time does not affect the score appreciably. However, the tester should be precise about the time-span since the statistical data and the norms are computed according to the time limits given in the instructions.

C. Applications of the Battery.

This battery is designed to be used as an aid in counseling and guidance for pupils in the 9th and 10th grades of academic high schools. The norms given in Table VII allow the calculation of the comparative level of the student with the national standard.

Knowing the levels of students is particularly important when dealing with youngsters in need of counseling and guidance, when, for example, they are on the threshold of dropout or when there is a questionable situation around graduating a certain pupil to a higher class. Class grades and the evaluation of teachers who have known the pupil over a year or two are, of course, consequential factors in such situations; however, an accurate measuring instrument standardized on national norms would serve to elucidate the picture of the problematic student.

It is often possible to prevent dropout from the educational framework by channelling the pupil to an educational environment more suited to his level. We emphasize here that this battery is not yet complete and is inadequate in its present form to function as a basis for the removal of a pupil from an academic high school to a non-academic one. Such a step involves the investigation of specific

abilities and an analysis of personality traits. The results of the battery provide only one factor of the many which have to be considered. Problematic cases should be dealt with by a counseling psychologist or by a vocational training centre.

From a psychological point of view, it is particularly interesting to note cases of significant discrepancies between test scores and school grades. Some of these discrepancies may be put down to random deviation stemming from inaccuracies in measurement. The results of any specific test might have been influenced by disturbed concentration or overtenseness. On the other hand, they might also have been influenced by a lucky guess or the appearance of well-known items in the test. The same applies to teacher evaluations, whose accuracy leaves much to be desired.

Significant discrepancies between the battery score and school grades usually indicate adjustment difficulties within the school framework, whether related to motivation or personality conflicts. These youngsters are in need of special attention from the school's counselling psychologist.

In addition to problematic children, the results of the battery may also be applied as a class scale.

Pupils may be graded according to the raw scores they achieved in the battery but it is easier to comprehend the significance of the scores when they are translated into norms (see Table VII). The norms bring the scores onto a scale of national standards.

The most accurate evaluation of the pupil's place in his class is based on the norm of his mean score in the battery and on the grade given by the teacher. The method for computing norms is detailed in part D.

The <u>class</u> as a whole can be graded on the basis of a national sample in the following way:

The mean score of the class is determined from the mean score of the pupils in each subtest. The data given in Tables V and VI

allow the teacher to locate the position of his class relative to the distribution of a national sample of class means. For example, if the mean score of a certain 9th grade is 13.3 in the mathematics subtest, this signifies that the class belongs to the 3rd quarter of the national sample in mathematics. A 10th grade with a mean score of 13.1 in literature would belong to the 2nd quarter of the national sample, and so on.

We have already noted that the score achieved in one test is not a reliable enough basis for the construction of individual profiles. The mean score of a class, however, is more stable since random deviations of measurement errors cancel themselves out and it is possible to compute the class level in each subtest separately.

TABLE V.

Interquartile ranges of the distribution of class means of each subtest for grade 9.

Subtest / class groups	Mathe-matics	Science	Reading Comp.-lit.	Reading Comp.-social studies	Social studies-general info.	English Language	Mean score in bat-tery
Up to first quartile	to 12.3	to 15.7	to 12.7	to 12.7	to 10.3	to 29.8	to 16.4
First quartile to median	12.4-13.1	15.8-17.0	12.8-13.7	12.8-13.9	10.4-12.2	29.9-33.6	16.5-17.1
Median to third quartile	13.2-14.9	17.1-19.6	13.8-14.3	14.0-14.2	12.3-14.0	33.7-35.6	17.2-18.3
Fourth quartile onwards	15.0-	19.7-	14.4-	14.3-	14.1-	35.7-	18.4-

TABLE VI. Interquartile ranges of the distribution of class means of each subtest for grade 10.

Subtest / class groups	Mathematics	Science	Reading Comp.-lit.	Reading Comp.-social studies	Social studies-general info.	English Language	Mean score in battery
Up to first quartile	to 13.2	to 16.6	to 13.0	to 13.0	to 12.7	to 35.5	to 18.1
First quartile to median	13.3-16.0	16.7-17.8	13.1-14.1	13.1-13.8	12.8-14.2	35.6-38.8	18.2-19.2
Median to third quartile	16.1-16.3	17.9-21.0	14.2-14.3	13.9-14.9	14.3-14.4	38.9-44.3	19.3-21.2
Fourth quartile onwards	16,4-	21.1-	14.4-	15.0-	14.5-	44.4-	21.3-

D. Distribution and Norms.

The distribution of scores of the battery in grades nine and ten are given in figure A.

The norms are based on a sample of classes, grade nine and grade ten, in accredited academic high schools. The sampling procedure was based on a stratification of schools by type of settlement and level of school (based on the results of matriculation examinations). The total number of cases was 370 from each grade. The Sample covered 21 classes. The data were collected during the final trimester of 1966.

The means of the distributions are 16.9 and 18.8 in grades nine and ten respectively. The standard deviations are 3.3 and 3.9 in grades nine and ten respectively.

The large overlap of the distributions shows that individual differences far exceed the differences between grade levels. Similar results can be seen by comparison of class means. For example, on the Social Science General Information subtest the means of the median

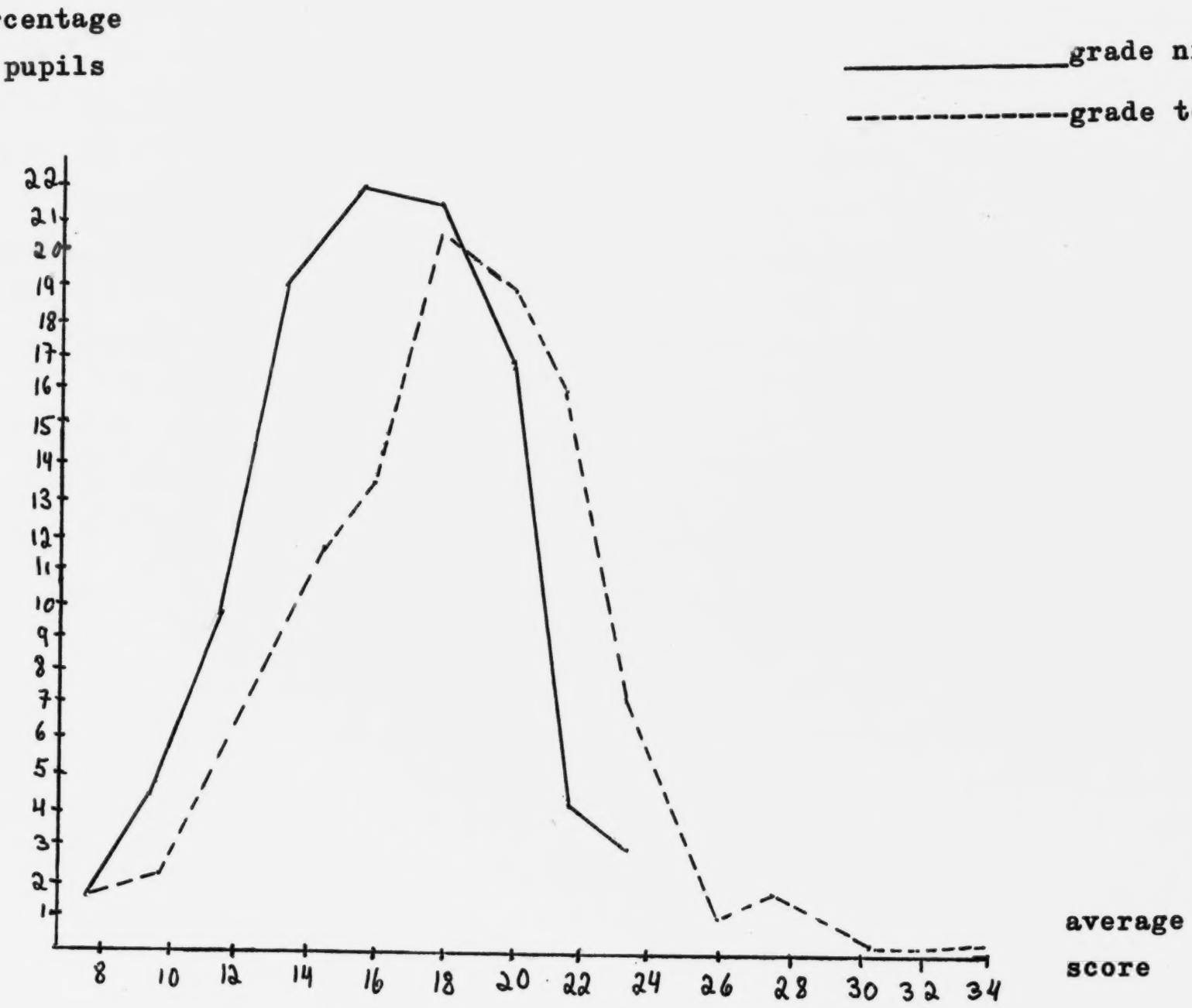
class of grade nine and grade ten are 12.3 and 14.3 respectively, the difference being 2.0, while the ranges of the class means of grade nine and of grade ten are 7.5 and 5.3 respectively. Similar results occur in the other subtests.

This is an expected outcome especially for a battery measuring general educational development, for attainment here depends on the background of the pupils, methods of instruction, extent of individual treatment, encouragement of extracurricular activities, and the selective policies of various schools rather than on progress during one year of high school. However, the available data are not sufficient for a thorough analysis of the factors involved.

For guidance purposes it is necessary to extend the norms in order to cover vocational schools and agricultural schools. This will be done in conjunction with the forthcoming validation and follow-up study.

Figure A

Distribution of scores (average of six subtests) in grade nine and in grade ten, in percentages



Mean of average score: grade nine - 16.9, grade ten - 18.8.
Standard deviation of scores: grade nine - 3.3, grade ten - 3.9.

TABLE VII.

Norms and percentile scores of the mean score.

Mean score achieved in the battery	GRADE 9		GRADE 10	
	equivalent standard in teacher's grade	cumulative % of pupils achieving this or lower score	equivalent standard in teacher's grade	cumulative % of pupils achieving this or lower score
31 plus			89 plus	100
30			84	99
29			84	99
28			84	99
27			82	98
26			82	98
25	87 plus	100	82	98
24	84	99	79	95
23	81	97	76	90
22	81	96	71	82
21	79	94	69	76
20	75	87	66	65
19	71	77	65	57
18	68	66	63	46
17	65	56	61	35
16	62	44	59	26
15	60	34	58	21
14	58	23	57	14
13	55	15	56	9
12	53	10	54	6
11	50	6	53	4
10	48	3	49	2
9	45	1	46	1
8 up to and incl.	45 up to and incl.	1	46 up to and incl.	0

The norms given in Table VII are based on a transformation of the battery scores equating them to the distribution of teachers' point grade average. This method was preferred here as it translates the scores into a scale easily comprehended by teachers. It should be noted that though the range of teachers' grades in each subject matter is 40-100, the range of the point grade average is contracted and extends from 45 to 87.

In the case of a pupil missing a test, or of the test being disqualified for technical reasons or by chance, the mean score of the 5 tests will be computed and the following points (Table VIII) will be added or subtracted from the resultant score (rounding to whole numbers will be done after the correction).

TABLE VIII.

Number of points necessary to add or subtract to scores of pupils participating in 5 subtests

Omitted subtest	procedure	9th grade	10th grade
Mathematics	subtract	.8	.8
Science	-	-	-
Reading Comprehension -literature-	subtract	.8	1.1
Reading Comprehension -social studies-	subtract	.8	1.0
Social Studies -general information-	subtract	1.0	1.2
English Language	add	3.2	4.1

The purpose of these corrections is to compensate for the differences in the test means. For example, a pupil who missed the English test will lose more points than a pupil who missed the literature test. The corrections provide the differences between the mean (of the total sample) in 6 tests and the means (of the total sample) in 5 tests.

The Table of norms is based on the means of 6 tests and if it is used to compute the mean in 5 tests a certain inaccuracy will result. The reliability of the mean of 5 tests is also affected to a certain extent.

Where 2 or more tests are omitted the battery cannot provide a basis for the construction of individual profiles.

E. Instructions for Administering the Battery.

1) The battery may be administered to the class as a whole or to individuals.

2) The test will be administered by the class teacher. The presence of one other person is desirable. The pupils should be spaced out sufficiently.

The administrator of the battery should arrange the technicalities in advance (proper and convenient seating order and space, sufficient supply of writing materials, etc.) since the test time is limited and technical hindrances during the test could disturb the time factor.

3) The tests will be given in two sittings with a few days' interval between them.

4) The administrator will address the following words of explanation to the pupils before they begin. The purpose is to reduce the tension of an examination and, at the same time, to arouse a serious attitude towards the battery. The explanation:

We are going to give 6 tests to the class, 3 today and 3 on... The tests cover specific knowledge in different subjects and general knowledge. You will come across questions on subjects which you have not studied or read about. This is because the tests have also been designed for classes with a different school curriculum from yours. However, you will find enough questions which you will have no difficulty in answering.

Not everybody will be able to finish the tests within the time allotted. If you finish before time is up you should check your answers.

After the tests are distributed we shall read the instructions together. Any questions you may have connected with the instructions will be answered. I will tell you when to open the questionnaire and you will all start together. If anyone has a question during the test he will raise his hand and I'll come to him. I'll answer only routine questions and will not answer questions connected with the material in the test. When I tell you to stop you will all put down your pens.

The first test is Science and after it has been distributed we'll read the instructions together.

5) The test administrator will distribute the questionnaires with reply sheets. The pupils will fill in their names on the reply sheets. The administrator will read the instructions for the first test aloud. He will be strict about seeing that the whole class starts and stops together.

6) Before starting the next tests the main points of the instructions will be repeated briefly as a reminder. It is also desirable to repeat the instructions before the second examination period.

7) The order of the tests and the time allowed for each one:

<u>First examination period</u>.

1) Social Studies - general information	15 minutes
2) Mathematics	50 minutes
3) Reading Comprehension - literature	30 minutes

A total of 1.35 minutes. With the explanation and distribution less than 2 hours.

The mathematics will be given straight after the social studies paper. Then it is worth having a 10 minute interval before the last test. Rough paper will be distributed for the mathematics test.

<u>Second examination period</u>.

4) Reading Comprehension - social studies	30 minutes
5) Science	30 minutes
6) English Language	30 minutes

A total of $1\frac{1}{2}$ hours. With the explanation and interval, less than 2 hours.

Science will come immediately after the reading comprehension test and then a 10 minute interval before the English paper. There is no reply sheet for the English test, answers are written on the questionnaire.

8) The tests will be marked according to a correct answer sheet which will be supplied to the teachers.

9) A pupil's score in each test is the number of correct answers. The mean score will be calculated on the basis of the total number of points scored in all 6 tests, divided by 6. Norms are given in Table VII.

F. REFERENCES

1. Bentwich, J.: Examinations and evaluation in secondary schools. Megamot, 1963.

2. Eson, M.E., Bernstein, R.J.: The curriculum and problems of evaluation. Megamot, 1958.

3. Guttman, E.L., Ortar, G.: Examination of validity of entrance examinations for candidates of the School of Medicine. The Institute for Social and Applied Research, Jerusalem.

4. Muhsam, H.V., Klinov-Mallul, R., Chanoch, G.: The supply of professional manpower from Israel. Educational Sources, Jerusalem, 1959.

5. Orr, L.: The reliability of the Israeli Bagrut examinations. Megamot, 1963.

6. Pilliner, A.E.G.: The Measurement of Educational Achievement in Israel, 1961.

7. Smilansky, M.: Outline for reform in secondary education. Megamot, 1961.